Segmenting Tumors Using a ML Model
DS4002 Case Study
Hudson Noyes

Background

AI-powered medical imaging is transforming healthcare by applying machine learning, especially deep learning, to automate critical tasks like segmentation, classification, and registration of medical images. These tools are particularly impactful in radiation oncology, where models assist in segmenting tumors and organs-at-risk (OARs) from treatment planning scans. This is vital for hard-to-identify cancers where manual segmentation is time-consuming and often inconsistent.

By improving accuracy and reducing clinical workload, AI models are standardizing care across patients and locations. With technologies like PyTorch, models are now approaching expert-level performance while operating much faster. This rapid progress signals immense opportunity for AI to scale access to high-quality care and enhance treatment outcomes globally.

The Deliverable

Participants will take our study a step further by training a custom PyTorch model to segment kidney tumors from a labeled dataset. While pretrained models have been explored, this challenge focuses on building one from scratch, offering hands-on experience in medical image analysis and model development.

We provide:

- A fully prepared dataset of annotated kidney tumor scans,
- Preprocessing scripts and data loaders,
- A technical guide titled "Training a Classifier" covering loss functions, optimizers, and training strategies.

Participants must apply a train-test split and avoid overfitting by using appropriate validation techniques and segmentation metrics (e.g., Dice score, IoU). Success will require solid understanding of machine learning fundamentals, model training, and Python programming. The final product should be a model that can accurately and efficiently segment kidney tumors, showcasing the participant’s technical ability and understanding of real-world AI in medicine.